DOCUMENT RESUME

ED 279 707 TM 870 094

AUTHOR Wolf, Richard M.
TITLE NAEP and International Comparisons.
PUB DATE 86
NOTE 20p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-093.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Testing; *Cross Cultural Studies; *Educational Assessment; Educational Policy; Educational Testing; Elementary Secondary Education; *Evaluation Needs; Financial Support; *International Cooperation; International Programs; National Surveys; Testing Problems; Testing Programs; Test Interpretation
IDENTIFIERS *International Evaluation Education Achievement; *National Assessment of Educational Progress

ABSTRACT

This paper examines a number of issues surrounding the addition of a dimension of international comparisons to the National Assessment of Educational Progress (NAEP). The basic position of the paper is that adding an international dimension to NAEP is certainly possible, but some problems will need to be overcome. Non-issues are identified and eliminated, including: (1) the organization of United States participation in the International Association for the Evaluation of Educational Achievement (IEA); and (2) response rates in IEA studies. Issues to be considered fall into three major categories: (1) financial and political, (2) technical, and (3) interpretation. Financial and political problems center around control and support for IEA projects. Another political problem relates to an IEA policy of not releasing any nation's data to another nation until an international report has been released. Technical issues include age and grade levels for testing, test construction, testing conditions, test organization, time of testing, and schedules of studies. When comparing nations, the results will need to be intepreted carefully. An appendix contains background materials about IEA. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEST COPY AVAILABLE

NAEP and International Comparisons

Richard M. Wolf

Columbia University

ED279707

Paper commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

1986

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. M. Wolf

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 094

NAEP and International Comparisons

Richard M. Wolf

Teachers College, Columbia University

Introduction

The purpose of this paper is to examine a number of issues surrounding the addition of a dimension of international comparisons to the National Assessment of Educational Progress (NAEP). The basic position that is taken throughout this paper is that adding an international dimension to NAEP is certainly possible, but some problems will need to be overcome. In addressing the overall problem, a number of issues will be identified and discussed. These issues fall into three major categories: (1) financial/political, (2) technical, and (3) interpretation.

The major (if not the sole) international organization for NAEP to join with in order to add an international dimension to national assessment is the International Association for the Evaluation of Educational Achievement (IEA). IEA has existed for over 25 years and has conducted a number of multi-national studies. The United States was one of the founding members of IEA and has been active in all aspects of the organization. Currently, IEA consists of over 40 member nations that are actively involved in a number of research projects. The attached guide, IEA: Activities, Institutions and People, presents background information about IEA.

The relationship between NAEP and IEA dates back to the late 1960's. At that time, IEA was seeking financial support from the U.S. Office of Education to support U.S. participation in a major survey of student achievement in six subject areas in a number of countries. The U.S. Office

of Education had recently committed itself to supporting NAEP and questioned whether supporting U.S. participation in IEA would be redundant. Accordingly, U.S. Office of Education personnel, in 1968, requested a paper detailing how IEA and NAEP differed. (See Appendix A.) After due consideration, the U.S. Office of Education supported the IEA enterprise as well as NAEP because it concluded that the two enterprises were sufficiently different.

In 1969, when the IEA Six Subject Survey was preparing to go into the field, it was felt that there might be problems in competing with the NAEP data collection. Accordingly, meetings were held with the Research Triangle Institute, the organization drawing the NAEP samples and overseeing the data collection, to develop ways of insuring that the two studies would stay out of each other's way. A cooperative solution was worked out and both enterprises accomplished their missions.

Since the early 1970's, there has been little contact between the two organizations. This has occurred mainly because each has gone its own way. Cooperation between NAEP and IEA could be reestablished quickly and satisfactorily. In fact, this seems to be happening. The Director of NAEP (LaPointe) is a recently added member of the US-IEA Advisory Council. He has attended one meeting of that group and is fully informed as to IEA's present and projected activities. In turn, IEA is committed to cooperating with both national and international organizations in promoting good educational research and evaluation.

Non-Issues

Before addressing the issues involved in adding a dimension of international comparisons to NAEP, it seems wise to dispose of some non-

issues. These are matters that may seem to pose problems but, in fact, do not. Dispatching them at an early stage should allow attention to be focussed on real issues.

The first non-issue involves the organization of IEA. The main governing body of IEA is its General Assembly. This group meets once a year and conducts the business of the organization. The General Assembly consists of one person from each member country. Currently, the National Center for the United States in IEA is Teachers College, Columbia University. This writer serves as the U.S. General Assembly Member. Discussions are proceeding on changing the U.S. National Center to the National Academy of Science (operating US-IEA as a panel of NAS). Presumably, a different person would be the U.S. General Assembly Member. Whatever happens is likely to have no effect on the issues involved in adding an international dimension to NAEP. IEA is open and cooperative with all organizations and this is not likely to change if the U.S. National Center is or is not part of the government. Recent history bears this out. Representatives of the Department of Education have attended and participated in IEA General Assembly Meetings regularly since 1982 and have had access to IEA documents, instruments, and results. NAEP personnel, if they wish, can have the same access. A simple letter or a phone call is all that is needed. Whether U.S. participation is governmental or non-governmental will not change this basic state of affairs. Thus, the organization of U.S. participation in IEA is a non-issue insofar as adding an international dimension to NAEP is concerned.

A second non-issue centers around the matter of response rates in IEA studies. Response rates in the U.S., in terms of the percentage of schools accepting invitations to participate in IEA studies, has been rather

low in two studies. Recent studies, notably writing and the second wave of the science study, have had much higher response rates. Improvements have been effected which should insure acceptably high response rates in future studies. What is also important is that response rates in other IEA countries are typically very high. Except for the U.S., Sweden, Australia and England, response rates in participating countries are all above 95%. Thus, one can be confident about results from other countries. Even in Sweden, Australia and England the response rates are at acceptable levels. Response rates in IEA studies are generally very high and even the U.S. is improving notably in this area. This means that one would be able to have confidence in results obtained in international comparisons. Thus, response rates seem to be a non-issue for the future.

Issues

While the matters of organization and response rates are non-issues, many issues still surround the addition of a dimension of international comparisons to NAEP. In this section, these issues will be identified and discussed under three major categories: (1) financial/political, (2) technical, and (3) interpretation.

Financial/Political

Historically, IEA has relied heavily on the United States for financial support. The U.S. government, foundations and other private organizations have not only supported U.S. participation in IEA studies, but also a good deal of the international costs of IEA studies. If the U.S. did not provide considerable support for IEA work, it is quite likely that IEA would soon go out of existence. In that case, there would be no basis for adding an

international dimension to NAEP since there would not be a group of countries doing studies using common instruments and procedures.

The costs of doing multi-national research are substantial. Instruments and procedures must be developed and tried out in a number of different countries, often representing rather differing cultures, in order to insure comparability. Simply importing instruments and procedures into one nation that were developed elsewhere does not work. IEA researchers have developed complex procedures for insuring that the quality of their work is as high as possible. Achieving this goal requires resources.

At times, the U.S. government has supported IEA work generously. At other times, U.S. support has been rather sporadic and at a low level. IEA cannot do high quality work without reasonable and dependable levels of support.

In past studies, the U.S. government has provided considerable support for IEA work. It has paid for much of the international development costs, full support of U.S. national costs, and a substantial part of international data analysis, meetings and report preparation. IEA is not likely to continue producing high quality international research without such continued support.

The U.S. government cannot and should not be expected to support multi-national research in the manner of a benign rich uncle, doling out money to an eager set of nephews and nieces. The U.S. government not only has a right but also a responsibility to see that its funds are spent wisely and well. Current discussions about the organization of U.S. participation in IEA as well as the establishment of a US-IEA Advisory Council that includes U.S. government representatives are aimed at insuring that the government has a say in helping to decide what work IEA will do, what studies the U.S.

will take part in, and how U.S. participation in IEA will be organized and managed. The problems clearly center around the issues of control and support. It is probably improper to state the issue as how much control will the U.S. government exercise for how much support. It is more complex than that. The U.S. government probably does not want to control and dominate an international research organization just for its own purposes. On the other hand, the U.S. government does have a legitimate interest in seeing that its informational needs are heard and some number of them are met.

Two minor organizational problems should be noted. IEA operates under a set of policies that are designed not only for the well being of the organization but also its member nations. Two policies that directly affect NAEP adding an international dimension to its work involve the release of IEA instruments and release of data sets. The first policy states that IEA does not release its instruments to other groups to use until after they have been used by IEA on its own data collection. The reason for this policy is that if IEA works to develop international instruments, it is entitled to use them first. This should not pose any great problem for NAEP use of IEA instruments. It is a matter of timing and should be resolved satisfactorily with joint planning.

The second policy is somewhat more of a problem. IEA has a policy to not release any nation's data to another nation until an international report has been released, a nation's own national report has been released, or a country has given permission for its data to be released. The reason for this policy is to insure that people in a country do not learn about their own students' performance from sources in other countries.

This policy could pose a problem. Suppose NAEP used a set of IEA test items in science. The items are administered in, say, 1988. NAEP processes

its data speedily and is prepared to issue results in early 1989. A comparison with Japan is desired. NAEP could not, in this example, publish any results for Japan until an international science report is issued in, say, 1991 or until Japan published its own national report in, say, 1990.

There is no easy solution to this problem. The integrity of each nation's data sets must be respected. The problem can probably be ameliorated with more resources. For example, if the Second IEA Math Study received timely support it might have completed its work much earlier. In any case, the problem needs to be recognized. It is not easily solved but can be dealt with to some degree through careful planning, including provision for adequate and timely support.

Technical Issues

Age/Grade of Testing. This is an easy issue to address. Solving it, however, is another matter. Simply put, the age or grade levels at which NAEP and IEA test do not usually match. There are both historical and rational reasons for this. NAEP tests students at ages 9, 13 and 17. IEA practice varies somewhat. In math, IEA tests at grades 8 and 12, roughly ages 13 and 17. However, in most other school subjects IEA tests at the grades where 10 and 14 year olds predominate as well as in the senior year of high school (which can range from age 16 to age 20 in some countries). The reason for heavy reliance on the grades in which 10 and 14 year olds predominate are: (1) age 10 is the last year in all IEA countries in which a single class teacher gives instruction in all major subjects, and (2) age 14 is the last year of official compulsory schooling in IEA member countries. Furthermore, IEA has baseline data on student achievement from the 1970's

at these age levels. Any attempt to use IEA instruments or selected items must address this issue, possibly through special equating studies.

Test Design. IEA achievement tests are designed according to an elaborate procedure that involves national curriculum analyses, assembly of an international test plan that builds on the national analyses, development of a test built according to the international test plan and several rounds of tryout of items in a number of different countries representing different languages and cultures. While arduous, the process is intended to produce test instruments that are clearly curriculum based and yield meaningful subscores and total scores. At worst, the tests are equally unfair to all participating countries.

In contrast, NAEP designs instruments that are presumably based on objectives formulated by groups largely outside of the educational enterprise and yield scores that are related to those objectives and not necessarily to a curriculum. Clearly, the orientation, procedures and resulting scores are different. If NAEP were to use IEA instruments, there would have to be an extremely careful and judicious selection of test items and exercises. This is not to say that the job can not be done, but it has to be done very carefully.

Test Conditions. Most IEA instruments are administered in a group setting with examinees reading and answering questions. In contrast, NAEP, I understand, uses an oral presentation of items to eliminate the reading burden. If so, this introduces a source of incomparability that is not easy to overcome.

IEA uses a separate answer sheet in the testing of students. If NAEP also uses a separate answer sheet, there is no problem.

IEA establishes fairly liberal time limits for all regular instruments so that students have the chance to at least attempt every item. If NAEP follows the same procedure, there is no problem.

Test Organization. IEA tests are developed so that, at each age or grade level, every student answers a fairly large common set of items. In addition, there may be additional sets of items at the end of the test that are rotated. In contrast, NAEP instruments do not employ a common set of items but use strictly rotated forms. (Technically, the term is matrix sampling.)

IEA follows its organizational plan because it seeks to study relationships between achievement and other variables of interest at the student level as well as at the class and school levels. NAEP, in contrast, has historically been interested in reporting the performance of particular subgroups in a population of 9, 13 or 17 year olds. (See Appendix A.)

There is no particular reason why NAEP cannot alter the organization of IEA instruments for its own purposes. IEA instruments are developed on the basis of an assumption of independence of items. Evidence gathered to date support that assumption. The only exception to this is in the area of reading comprehension and literature where a set of items is generated for each reading passage. In such cases, there is probably some item dependence. However, it is not likely that anyone would break up a reading passage and its set of test items.

Time of Testing. IEA studies typically collect data at the end of the school year. There are good reasons for this. If IEA is studying student achievement in a one year biology course, end of year testing is the only sensible course of action. The same is true of subjects such as pre-calculus math. Of course, the end of the school year varies enormously in IEA

countries. In some cases it's June while in others it may be December or March. (IEA has southern hemisphere countries where the end of the school year and beginning of summer comes in December.) When NAEP tests in the school year is not known to this writer. If NAEP's time of testing in the school year differs from IEA's, some work may be required to obtain comparable estimates for achievement between the U.S. and other countries. This is not an insurmountable problem.

Schedule of Studies. NAEP schedules testing in various school subjects on a regular multi-year basis (although it has departed from its schedule on occasion and changed its schedule at other times). IEA, in contrast, has been highly irregular, even sporadic, in the scheduling and conduct of studies. Partly this is a function of support. When money is available, IEA is active. It is hoped that IEA will move to a more regular schedule of testing in basic school subjects. A proposal along these lines is scheduled to be considered at the upcoming IEA General Assembly meeting in August, 1987. If accepted, a regular cycle of testing will be followed and organizations such as NAEP can plan their work on the basis of a known IEA schedule.

Interpretation

Perhaps the largest problems to be addressed if an international dimension is added to NAEP deal with the interpretation of results. NAEP can easily obtain and use IEA test items and entire instruments if it wishes. Furthermore, the administration of IEA instruments can be tied in with NAEP studies. Also, NAEP can gain access to data sets from various

countries. All this can be arranged with relatively little difficulty. (IEA data sets are available to bona fide researchers anywhere in the world.)

Given all of the above, one will need to interpret results very carefully. It will be easy to report a mean or percentage difference between the U.S. and some other nation. It will also be relatively easy to evaluate the statistical significance of such a difference. (Since IEA and NAEP studies use very big samples almost any difference is apt to be statistically significant, even with complex sampling designs.) What will be difficult is making sense out of a difference. This will require a good deal of knowledge about the U.S. educational system and the educational system of each nation that one wants to compare U.S. performance with. It is a matter that IEA personnel have struggled with for years. In one respect, IEA has a great advantage in this area. Since IEA is composed of a number of participating nations, members from participating countries are simply asked to explain particular results. Often, IEA members can produce explanations for particular results because they know their own educational and social systems so well. At other times, however, even IEA members are hard pressed to explain a particular result for their country. Interpretation of results is and will probably continue to be a problem.

Some examples are perhaps in order. Consider the issue of apparent U.S. decline in relative performance with increasing age. In the very first IEA study in math in 1964 it was noted that younger students in the U.S. (13 year olds) performed relatively better (compared to students in other nations) than 15 and 17 year olds. In fact, U.S. high school seniors scored just about the lowest of all participating countries. This result occasioned a front page article in the New York Times and, two days later, a lead

editorial in the New York Times titled, "The Math Gap." It was a time for hysteria.

It was only later that people realized that while the high school senior group in the U.S. contained about 70% of the 17-18 year old age cohort, the high school senior group in other countries was as low as 10% of the age cohort. In fact, no other country came close to the U.S. in degree of retentivity. When adjustments were made for retentivity (really selectivity), the results changed rather dramatically. Rather than being dead last, the U.S. was in the middle of the pack of countries and, in fact, had the highest "yield" of math competence of any country. (See T. Husen (ed.) International Study of Achievement in Mathematics. Stockholm: Almqvist and Wiksell, 1967). Since then IEA researchers have been quite sensitive about taking system retentivity into account in interpreting results.

Now consider the problem of interpreting U.S.-Japanese differences in math performance at the eighth grade level. The recent Department of Education publication, What Works, contained a bar graph showing various country's performance in math. Since this is a grade when schooling is compulsory, the U.S.-Japanese difference cannot be interpreted in terms of differential retention or selection. So why is Japanese performance so high and U.S. performance so low on this set of math items?

There is no easy answer. However, consider the following facts about the two educational systems. First, math achievement is highly prized in Japan--more so than in the U.S. Second, the Japanese school year consists of 220 full day equivalents of instruction compared to the 180 day school year in the U.S. Third, an enormous number of Japanese students attend the Juky, the coaching schools, where a great deal of additional instruction and

practice is given, mainly in math. Conservative figuring shows that Japanese students receive at least 50% more instruction and practice in math in a school year than U.S. students. Given just these facts, one could legitimately inquire why the difference between U.S. and Japanese students is not greater than it is! The point is that simply trying to interpret results from different countries is fraught with difficulties.

Summary

This paper has attempted to identify and address a number of issues related to adding an international dimension to NAEP. It has also attempted to identify and dispose of some non-issues. The issues fall under the major categories of financial/political, technical and interpretation. While the issues are real and, in some cases, formidable, they are not insurmountable. If there is a strong commitment to adding an international dimension to the work of NAEP and provision for resources to do the job, it can be done. It will require work. IEA has had a longstanding policy to work cooperatively with other researchers and research organizations. There is ample reason to believe that IEA will cooperate with U.S. organizations in adding an international dimension to NAEP's work if that is what is wanted for NAEP.

APPENDIX A

Differences Between the International Educational Achievement Project (IEA) and the National Assessment of Education (NAE)

The purpose of this paper is to set forth differences between the International Educational Achievement Project (IEA) and the U.S. National Assessment of Education (NAE). It is hoped that a statement of the differences between these two enterprises will be of some use in the separate and joint consideration of both. The differences are listed under the heading of purpose, method and consequences.

Purpose

1. IEA seeks to determine the functional relationships between inputs and outcomes of the educational process in a wide variety of settings, i.e., 19 countries. Inputs include school organization, teaching and administrative practices, and teacher, pupil, home and community variables. Outcomes include achievement in various subject areas and student attitudes.

2. NAE seeks to describe the level of educational outcomes in various subject areas for the cross classifications of geographic region, sex, community size, race and socio-economic status for four age groups.

Method

3. IEA has predictor and criterion variables; NAE has only criterion variables.

4. IEA is analytical and normative; NAE is normative.

5. In the IEA study, pupils and schools are units of analysis. Representative samples of 10-year olds, 14-year olds and high school seniors will be tested in the 19 participating countries. In the NAE study, results will be presented for two sexes, four geographic regions, four community sites, two races and two-socio-economic levels for 9, 13 and 17-year olds and young adults between the ages of 26 and 35.

6. IEA will collect data on student achievement, attitudes, ability, home background, teachers, schools and community; NAE will collect data on student achievement.

7. Test construction activities differ markedly between the two enterprises. In the IEA project, each country convened a committee of experts for each subject area and prepared a national paper on the objectives and content of instruction for the subject area after having carried out an extensive content analysis of textbooks, external examinations and curriculum guides. The national papers for each subject area were submitted to an international committee in that subject area. On the basis of these national papers, the international committee prepared a set of specifications for international tests in the subject area. These were sent to the national committees for review and some modifications were made. Items were selected and constructed which met the test specifications and then pretested in the various countries. Three times as many items as would eventually be required were tried out. Criteria for inclusion of items in the final tests included both logical and statistical characteristics of the items.

In the NAE project, objectives and prototype exercises for each subject area were origionally formulated by testing agencies who worked with scholars, teachers and curriculum specialists. The objectives, and

prototype exercises which emerged were reviewed by lay panels and a final set were agreed upon. Contractors were then asked to construct a large and representative set of exercises for each subject area for the four age groups. These were then pretested and final selections of exercises were made.

8. Procedures for analyzing the data and reporting will differ sharply in the two projects. For the IEA study, descriptive statistics (means, standard deviations, frequency distributions and correlations) will be reported. In addition, extensive use of multiple regression analysis will be made to determine the functional relationships between the input and output variables. In addition, a principal components analysis of the subscores in each subject area will be carried out for each country. Loadings on the first principal factor will be used to weight the subscores so as to best represent the results for each country. In the NAE project, the percent of individuals attaining each objective in each of the units of analysis will be reported.

Consequences

9. Besides presenting normative data for each country as well as for the 19 countries, the IEA project hopes to inform policy making in the various countries by identifying those inputs of educational policy, school organization and administration, teachers and teaching which affect achievement. Furthermore, besides studying the effect on achievement of each variables singly, provision has been made for studying the joint effects of variables. In this way, it will be possible to identify combinations of variables in an educational system which have consequences for achievement rather than to study an aggregation of policies and practices in isolation from one another. The variety of educational policies and

practices in the 19 participating countries represents a range of policy alternatives far greater than currently available to U.S. educators.

10. The NAE project will present descriptive data for each objective for each of its units of analysis. The typical expression of results will be the percent of individuals in each of the units which have attained an objective. Since testing is to be carried out in a three-year cycle, it will be possible to ascertain changes in achievement over a four-year period. However, if there are changes in achievement, it will not be possible to identify the reasons for such changes.

Richard M. Wolf
Teachers College
Columbia University

References

1. C. Hightower, How Much Are Students Learning: Plans for a National Assessment of Education (Ann Arbor, Mich.: Committee on Assessing the Progress of Education, 1968).

2. J.A. Sonquist and J.N. Morgan, The Detection of Interaction Effects (Ann Arbor, Mich.: Institute for Social Research, 1964).

3. Various IEA documents.